Impact of Halogen Bonds on Protein-Peptide Binding and Protein Structural Stability Revealed by Computational Approaches

Jintian Li[1,2], Liping Zhou[1,2], Zijian Han[1,2], Leyun Wu[1,2], Jianfang Zhang[1], Weiliang Zhu[1,2,*], Zhijian Xu[1,2,*]

[1]State Key Laboratory of Drug Research; Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, 201203, China

[2]School of Pharmacy, University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing, 100049, China

[*]Corresponding Authors

E-mail: zjxu@simm.ac.cn (Z.X.), wlzhu@simm.ac.cn (W.Z.).

Abstract

Halogen bonds (XBs) are essential non-covalent interactions in molecular recognition and drug design. Current studies on XBs in drug design mainly focus on the interactions between halogenated ligands and target proteins, lacking a systematic study on naturally existing and artificially prepared halogenated residue XBs (hr_XBs) and their characteristics. Here, we conducted a computational study on the potential hr_XBs in proteins/peptides using database searching, quantum mechanics calculations, and molecular dynamics simulations. XBs at protein-peptide interaction interfaces are found to enhance their binding affinity. Additionally, the formation of intramolecular XBs (intra_XBs) within proteins may significantly contribute to the structural stability of structurally flexible proteins, while having a minor impact on proteins with inherently high structural rigidity. Impressively, introducing halogens without the formation of intra_XBs may lead to a decrease in protein structural stability. This study enriches our comprehension of the roles and effects of halogenated residue XBs in biological systems.

Key Words: Halogen bonds; Binding affinity; Protein structural stability; QM/MM; MD

1. Introduction

In recent decades, it was confirmed that halogen atoms have the capabilities similar to hydrogen atoms when interacting with Lewis bases, which was called halogen bond (XB)[1]. The anisotropic charge distribution surrounding covalently bonded halogen atoms gives rise to a positively charged electrostatic region recognized as the σ-hole[2-4], which forms along the extension of C–X bonds (X = Cl, Br, I) and interacts attractively with the nucleophile[5-6]. The strength of the XB relies on the size of the σ-hole, which corresponds to the trend of halogen polarizability [Cl < Br < I][7]. Additionally, as the strength of the XB increases, their geometric shapes become more ideal (shorter bond lengths with angles closer to linear arrangement). Due to their critical role in the structure and function of molecular interactions in biological systems, XBs have attracted more and more interest in rational drug design[8-12].

In biological systems, the majority of research about XBs concentrate on the interactions between small organohalogens and proteins. A comprehensive analysis of structural data available in the Protein Data Bank (PDB) has illuminated potential interaction partners for halogenated ligands in the proteins[8]. In ligand-protein complexes, common XB donors are organohalogens, whereas the dominant acceptors consist of carbonyl oxygens in the protein backbone[8, 13-15]. The strength of the organohalogen XB can be additionally modulated by the electron-withdrawing capacity of substituents covalently bonded to a given scaffold, and the core scaffold that the halogen is attached to also play a role[10, 13, 16]. Numerous pharmaceutical ligands incorporate halogen atoms into their molecular structures, aiming to optimize the membrane permeability and metabolic stability of small molecules[17]. Based on a comprehensive database survey of Thomson Reuters Pharma and ZINC, Xu et al. noted that around one-third to one-quarter of compounds at different stages of drug discovery or launch are organohalogens, and halogenation may significantly contribute to improving protein–ligand binding affinity as well as ligand ADME/T properties[18].

When it comes to proteins/peptides, it's reported that naturally existing halogenated amino acids have an impact on both the structure and function of these

proteins. For instance, halogenated tyrosine and tryptophan in marine organisms have been found to promote protein cross-linking[19], while chlorotyrosine and bromotyrosine produced through oxidative halogenation are biomarkers for human asthma[20]. In addition, there are artificially modified halogenated proteins/peptides. It has been reported that the introduction of halogenated residue XBs can modify enzyme thermal stability, activity, and substrate selectivity [21-23]. Furthermore, the work of the Máté Erdélyi group suggests that the introduction of XBs can alter local protein folding[24].

To the best of our knowledge, no systematic study has been performed on the roles of halogenated residue XBs (hereinafter hr_XBs) in proteins/peptides. In this study, we conducted an analysis on the hr_XBs in proteins/peptides with PDB[25], followed by quantum mechanics (QM) calculations or molecular dynamics (MD) simulations, to explore the roles of hr_XBs. The results show that the intermolecular halogenated residue XBs between proteins and peptides can enhance their binding affinity. In addition, halogens within proteins have diverse impacts on protein structural stability, depending on both the formation of intramolecular halogenated residue XBs and the protein's inherent structure features. The results could be useful for researchers to utilize hr_XBs in halogenated peptide drug design and protein engineering.

2. Methods

2.1 Structure Survey for Halogenated Protein

The PDB [25] database (September 13, 2022 release) was used to explore the hr_XBs in proteins/peptides. Considering the long-range nature of electrostatic interactions and the fact that XBs are underestimated in PDB structures[26-29], both the typical [18]and looser geometry criteria of XBs were applied for detecting (potential) hr_XBs. The distance and angle of the looser criteria were defined to be 1 Å longer than the sum of van der Waals radii (Table S1) and larger than 120° ($\alpha < 70°$ and $\theta > 110°$ for the C−X···π type of XBs), respectively.

2.2 QM/MM Optimization

QM/MM was performed to optimize the proteins/peptides systems using the two-layer ONIOM (our own N-layered integrated molecular orbital and molecular mechanics) model[30] as implemented in Gaussian 16[31]. The key residues for proteins/peptides involved in hr_XBs were included in the QM region, whereas the rest were contained in the MM region. The QM region was described at the M06-2X/6-31g(d) level[32,33] for all atoms (LanL2DZ basis set[34] in the case of iodine) and the MM region was described by the AMBER parm96 force field[35].

2.3 Halogen Bonding Analysis

We conducted a Natural Bond Orbital (NBO) analysis to investigate the intermolecular electron transfer during the formation of hr_XBs. Additionally, the region, size and species of the weak interaction of hr_XBs between the proteins and peptides were studied by the Independent Gradient Model (IGM) analysis[36,37]. The analyses of molecular electrostatic potentials (MEPs) for halogenated residues, NBO and IGM were undertaken by the Multiwfn program[38], based on the obtained wave functions at M06-2X/6-311+G(d,p) level[39]. As for proteins, the adaptive Poisson-Boltzmann equation solver (APBS)[40] was used to compute the MEPs. All the analyses mentioned above were conducted based on QM/MM optimized structures.

2.4 Calculation of the Binding Energy

In order to evaluate the binding strength of halogenated and non-halogenated peptides to proteins, single point energy was calculated at M06-2X/6-311+G(d,p) level (LanL2DZ basis set[34] in the case of iodine) and the binding energy between peptides and proteins were then assessed by eq 1.

$$\Delta E = E_{com} - E_{res^{pep}} - E_{res^{pro}} + BSSE \qquad \text{(eq 1)}$$

where ΔE is the binding energy, E_{com} is the energy of the whole complex in the QM region, $E_{res^{pep}}$ and $E_{res^{pro}}$ are the energies of the key residues in peptide and protein, respectively, the basis set superposition error (BSSE) are considered in the binding energies calculation[41].

2.5 Extra Point (EP) Model

To account for intramolecular halogenated residue XBs in MD simulation, halogenated residues in proteins were parametrized using an EP model[42-43]. The structural optimization and electrostatic potential (ESP) calculation of halogenated residues were carried out at the HF/6-31G(d) level for all atoms (LanL2DZ basis set[34] in the case of iodine). An EP was then introduced to attach to the halogen atom along the axis of the C-X bond at a distance ranging from 1.00 to 2.50 Å with an increment of 0.01 Å, maintaining a fixed C–X···EP angle of 180°. The partial atomic charges were subsequently obtained by applying the Restrained ElectroStatic Potential (RESP) method using the Multiwfn software[38]. The selected X···EP distance was determined based on the best fit to the QM electrostatic potential, corresponding to the minimum relative root mean squared error (RRMSE) values. In addition to the X···EP distance, we employed EP parameters similar to those reported in previous studies (Table S2)[42].

2.6 MD Settings

Each protein system was solvated into a periodic TIP3P water box, extending 10 Å from the solute, and was neutralized by placing a rational number of counterions of Na^+ or Cl^-. The topology of protein system was generated using amber ff14SB force fields. About 8000 steps of minimization including 6000 steps of steepest descent minimization and 2000 steps of conjugate gradient minimization were performed to remove bad contacts during the energy minimization phase. The equilibration phase was then carried out separately in the NVT and NPT ensembles, each for 20 ps. Finally, the 150 ns MD production simulation was performed at 300 K, 1 bar, using a Langevin thermostat[44] and a Nosé–Hoover Langevin barostat[45-46] to control the temperature and pressure. Bonds involving hydrogen atoms were fixed by the SHAKE algorithm[47]. The cutoff distance applied for van der Waals interactions was 12 Å. The particle mesh Ewald method (PME) was used to deal with long-range electrostatic interactions. All MD simulations were performed using the GROMACS software package, version 2019.6[48]. Three simulation replicates were run for each

system.

2.7 MD Analysis

The root-mean-square deviation (RMSD) values were calculated for heavy atoms using initial structure as the reference. During the equilibrium period, the analysis of the root-mean-square fluctuation (RMSF) for the combined 3 trajectories was performed to evaluate the flexibility of the proteins. The radius of gyration (Rg) was calculated as a measure of the size and compactness for proteins. To analyze the effect of halogenation on protein structural stability, we used the software CONtact Analysis (CONAN)[49] to calculate the local interaction time (LIT) for individual protein residues of the combined trajectories. MDAnalysis[50-51] was used to calculate the occurrence of intramolecular halogenated residue XBs from the trajectories.

3. Results And Discussion

3.1 PDB Survey

Among the 195,093 structures in PDB, there are a total of 423 structures of proteins/peptides containing 2430 halogen atoms, 949 of which are heavy halogen atoms (Cl, Br and I). With typical XB criteria, 84 heavy halogen atoms (8.9%) were found to form 101 hr_XBs. However, if the looser criteria were applied, 295 heavy halogen atoms (31.1%) may form 532 potential hr_XBs, indicating that the hr_XBs might be highly underestimated in PDB structures. Among the 532 potential hr_XBs, 350 intermolecular halogenated residue XBs were found between the proteins and the halogenated peptides (hereinafter inter_XBs), while 182 intramolecular halogenated residue XBs were identified within the proteins (hereinafter intra_XBs). The distribution of halogen atoms and potential hr_XB in the 949 heavy halogen atoms are shown in Figure 1a. Compared with chlorine (35.6%) and bromine (21.0%), iodine (43.4%) is the most popular halogen atom in proteins/peptides. As shown in Figure 1b, considering previous literature[52], both organohalogen XBs and hr_XBs demonstrate a preference for sulfur atom as acceptor. However, for hr_XBs, side chain atoms

(nitrogen and oxygen) exhibit a higher propensity to act as acceptors compared to the main chain atoms, and this trend is not pronounced in organohalogen XBs[52], which may due to the preference of atom types around halogens in proteins/peptides (see Figure S1).

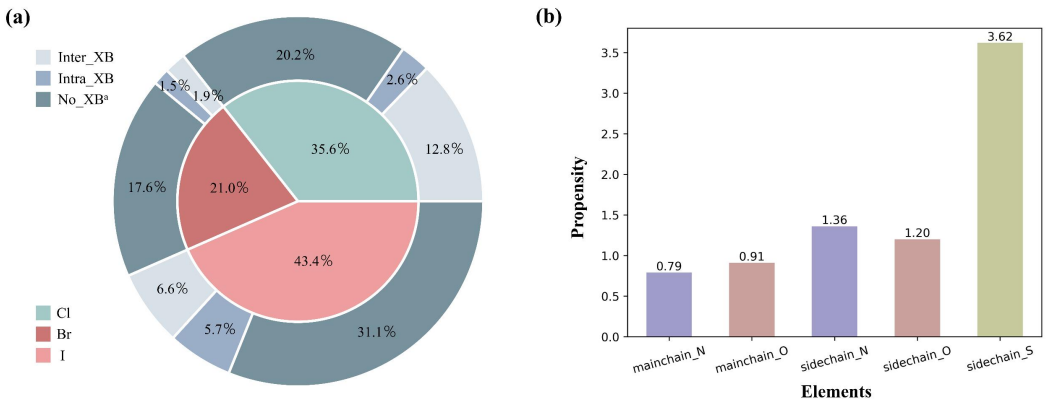

Figure 1 Statistical results of halogenated proteins/peptides. (a) Distribution of halogen atoms (excluding F) and potential hr_XB formations. (b) The elements propensity for hr_XBs (X···Y type). Propensity is defined as the relative frequency of each element in hr_XBs divided by its relative background frequency in the refined dataset (Identical hr_XBs occurring in the same system have been removed).

[a] No halogenated residue XB formed

3.2 The Strength of Intermolecular halogenated residue XBs (inter_XBs)

In order to explore the inter_XBs between proteins and peptides, we selected four systems for further QM/MM calculation. As evidenced by the results presented in Table 1, the optimized structures conform to the typical geometry criteria for XBs and have a better linear contact than their PDB structures (the geometries of the Cl system 1OKW are provided in Table S3). The geometries of the protein−peptide complexes after QM/MM optimization are shown in Figure 2a. It can be found that the XB regions change greatly after QM/MM optimization, which is different from the optimization results of halogenated ligand-protein systems[52]. However, the overall structural changes in the complex (including host protein and peptide) after optimization are minimal (Figure S2a), suggesting that the significant changes in XB

region might be caused by the relatively high flexibility of halogenated residues side chains. Furthermore, our research reveals that most of the non-XB interactions (HB, ionic, cation-π interaction, etc.) observed in crystal structures are preserved in optimized structures, with additional non-XB interactions formed after optimization (Figure S2b). Studies on the optimized structures of halogenated and non-halogenated complex suggest that the non-XB interactions may positively influence the XB region, while halogens still remain the primary contributors to the XB formation (Figure S3). The results of QM/MM optimization indicate that the inter_XB present at the protein-peptide interface in PDB structure can be well recovered (The details of XBs region for the systems calculated here are provided in Table S8).

Table 1 Protein-peptide complexes selected for QM/MM optimization

PDB ID	XB mode	Crystal		QM/MM	
		d(X···Y), Å	∠(C−X···Y),deg	d(X···Y), Å	∠(C−X···Y),deg
5ITF[a]	C−Br···O	4.37	146.4	3.03	170.6
1CZI	C−I···O	3.94	157.1	2.96	157.6
1GA4	C−I···O	3.47	104.5	2.99	150.2

[a] Calculation results for chains A and E (cAE) of the 5ITF system. The results for chains B and E (cBE) and chains D and F (cDF) are provided in Table S6.

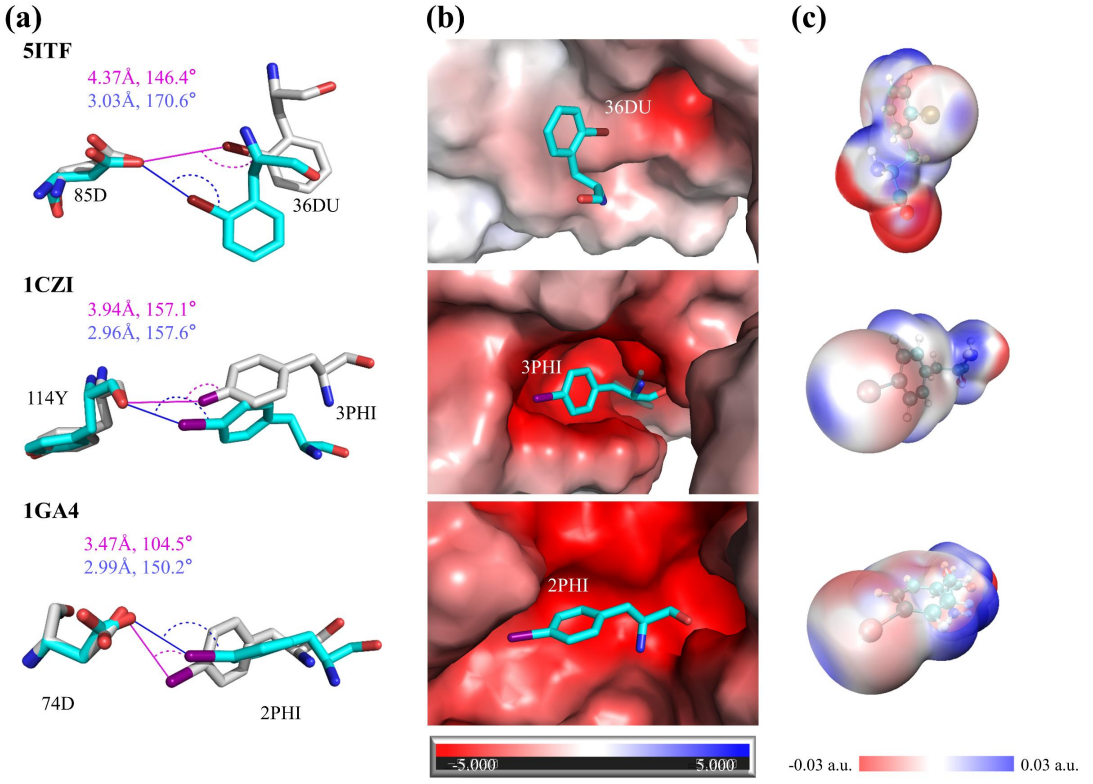

Figure 2 Structural characteristics and MEP of the XB region. (a) Structure details of the 3 systems before and after QM/MM optimization. The PDB structures are shown in gray, while the QM/MM optimized structures are illustrated in cyan. Interaction distances (Å) and angles (deg) for the PDB structures are presented in magenta, contrasting with the blue annotations for the optimized structures. (b) The MEP of the protein region located around the halogenated residues. (c) The MEP maps of the halogenated amino acid residues.

The MEP maps of the residues involved in the inter_XBs are shown in Figures 2b and 2c. The red areas correspond to regions with negative potentials, indicating the nucleophilic sites, while the positive potentials regions described by the blue areas, indicating the electrophilic sites. The observation of positive electrostatic potential regions, known as σ-holes, on the molecular surface of halogen atoms in all studied halogenated residues, along with the presence of negative electrostatic potential regions on the surface of the oxygen atoms of the XB acceptors, demonstrates the favorable electrostatic complementarity between the halogen atoms and the acceptor

atoms.

NBO analysis was conducted to explore the intermolecular electron transfer involved in the formation of inter_XBs. Table 2 presents the second-order perturbation energy (ΔE(2)) values between the inter_XB donors and acceptors in 3 systems. The NBO analysis results show that electrons transfer from the inter_XB acceptor's lone pair electron orbital of the oxygen atom to the donor's antibonding orbital of the carbon–halogen bond. Such an electron transfer phenomenon is consistent with the basic profile of XBs that electron always transfer from the XB acceptors to the XB donors.[53]

Table 2 Natural bond orbitals and the E(2) results

PDB ID	$E(2)_{sum}$[a]	E(2)[b]	Electron donor orbital	Electron acceptor orbital
5ITF	2.86	1.33	LP(1)O	BD*(1)C−Br
		0.51	LP(2)O	
		1.02	LP(3)O	
1CZI	2.38	2.18	LP(1)O	BD*(1)C−I
		0.20	LP(2)O	
1GA4	4.40	1.79	LP(1)O	BD*(1)C−I
		0.99	LP(2)O	
		1.62	LP(3)O	

[a] The sum of second-order perturbation energies

[b] Second-order perturbation energy

IGM analysis was carried out to explore the nature and strength of the weak interaction of the inter_XBs in the 3 complex systems. The type of noncovalent interaction can be characterized by the sign of the second largest eigenvalue of the electron density Hessian matrix (sign (λ2)). A negative value of sign (λ2) × ρ indicates an attractive interaction, while a positive value suggests a repulsive interaction. A value of zero corresponds to a van der Waals interaction. The IGM isosurfaces for 3

selected systems are displayed in Figure 3. The blue region in the center of the isosurface indicates the strong inter_XB interaction formed between the halogen atom and the acceptor atom.

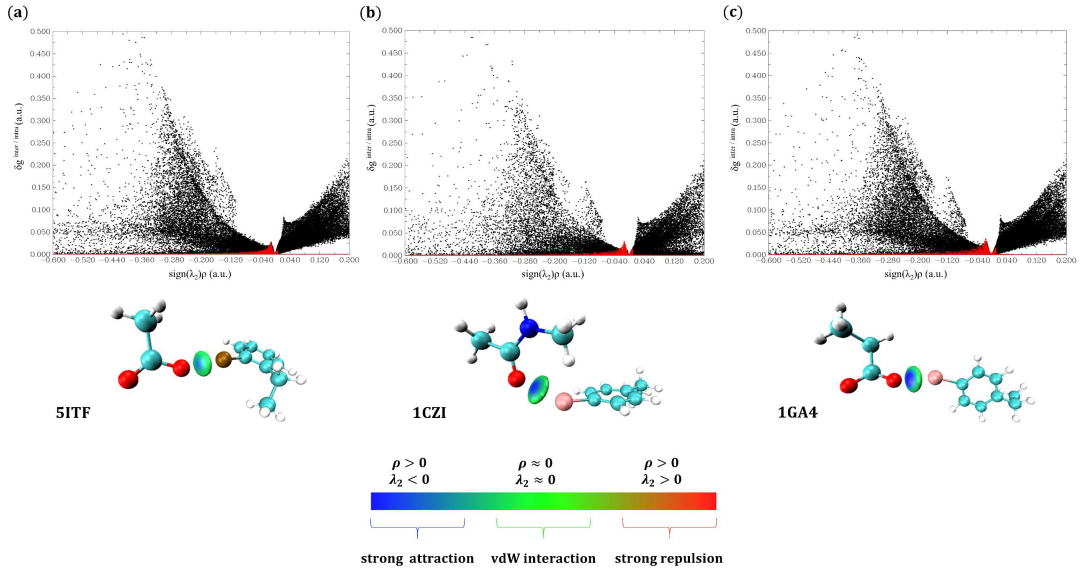

Figure 3 The IGM analysis and sign(λ_2)ρ colored isosurfaces of δg^{inter} = 0.01 a.u. for the selected systems (a) 5ITF, (b) 1CZI and (c) 1GA4 with ranging from -0.05 to 0.05 au. The regions participating in inter_XBs are displayed in a ball-and-stick format, where bromine atoms are rendered in brown and iodine in pink. The repulsive interaction regions are visualized in red, the weak interaction regions in green, and the strong attraction regions in blue.

The analyses presented above, considering both the geometric and quantum chemical features of the XBs, indicate the formation of inter_XB interactions between proteins and peptides. Furthermore, we calculated the binding energy for these inter_XBs to investigate whether halogenation can enhance the binding affinity. For each system, the halogen atom (X) in the halogenated peptide was replaced with hydrogen atom (H), QM/MM optimization was performed on it, and the energy difference (ΔΔE) was calculated. As shown in Table 3, the ΔΔE values calculated in vacuum for all systems are negative, indicating that the presence of inter_XB enhances the protein−peptide binding interactions (the results calculated in

chloroform are shown in Table S4).

Table 3 The binding energies of optimized structures for 3 systems in vacuum (kcal/mol)

System	ΔE_X[a]	ΔE_H[b]	$\Delta\Delta E$[c]
5ITF[d]	-5.39	-1.92	-3.47
1CZI	-2.33	-1.42	-0.91
1GA4	-4.21	-0.81	-3.40

[a] Binding energy of protein and halogenated peptide. The energies calculated for the crystal structures are shown in Table S5.

[b] Binding energy of protein and non-halogenated peptide.

[c] For each system: $\Delta\Delta E = \Delta E_X - \Delta E_H$.

[d] Calculation results for chains A and E (cAE) of the 5ITF system. The results for chains B and E (cBE) and chains D and F (cDF) are provided in Table S6.

3.3 Intramolecular Halogenated Residue XBs (intra_XBs) and Its Effect on the Structural Stability of Halogenated Protein

To implement the EP model in MD simulations, it is necessary to extract the optimal X···EP distances from two halogenated residues (4BF and IYR), which can best reproduce the ESP calculated by QM. As shown in Figures 4a and 4c, the optimal X···EP distances for Br (4BF) and I (IYR) are determined to be 1.37 Å and 1.73 Å, respectively, based on the minimum RRMSE values. Under these optimal X···EP distances, the EP model accurately reproduces the anisotropic nature of the ESP around the halogen atoms, with a σ-hole well evident on the halogen atoms (Figures 4b and 4d).

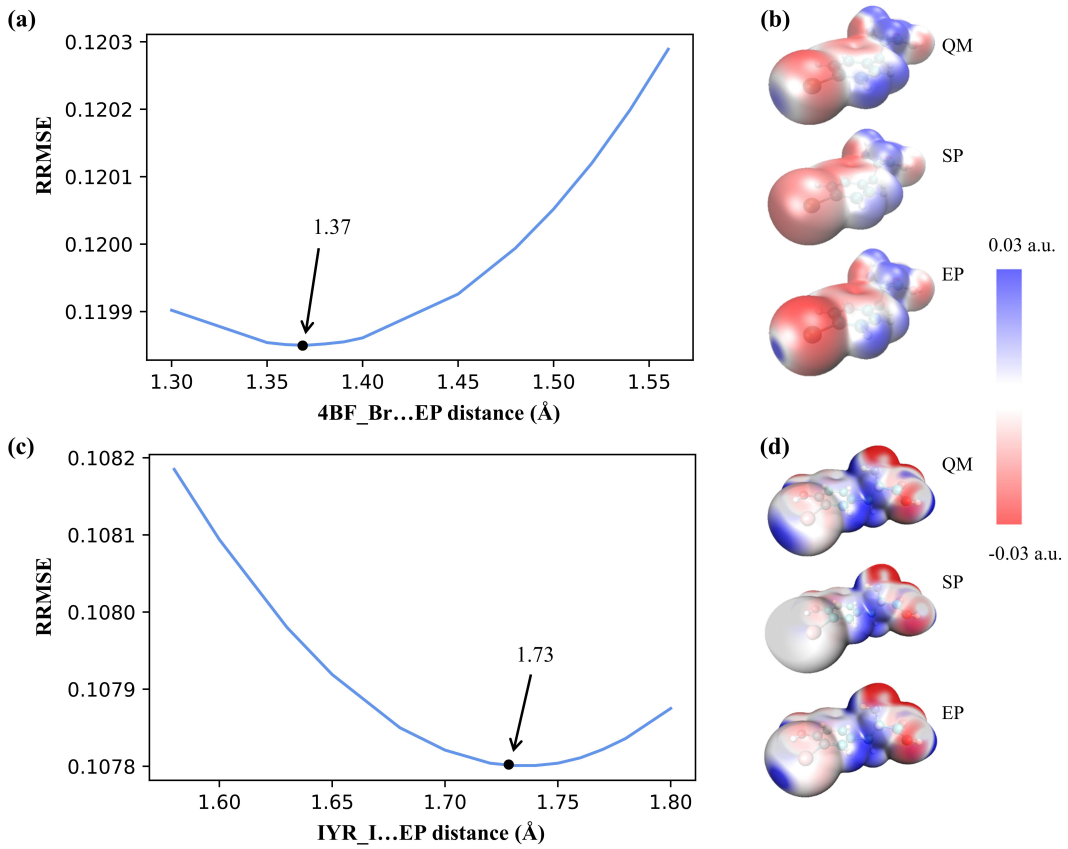

Figure 4 Distances of EP and MEP of halogenated residues (a, c) The X…EP distance fitting results of 4BF and IYR. (b, d) Electrostatic potential mapped onto the 0.001 a.u. electron density isosurface of 4BF and IYR using various methods (SP: single point charges).

Using the EP parameters described above, MD simulations were performed 150 ns each for the 3 systems (2ZXV, 3VN3, and 5KI8) to illustrate the structural stability upon halogenation. Notably, in the halogenated 2ZXV system (2ZXV_I), the average Rg measures 16.41 Å, while in the non-halogenated counterpart (2ZXV_noX), it registers at 16.88 Å (Figure 5), suggesting that 2ZXV_I is more stable than 2ZXV_noX. For the 3VN3 system, analysis of Rg indicates minimal differences between 3VN3_noX (16.81 Å) and the halogenated one (16.79 Å). In the 5KI8 system, 5KI8_Br has an average Rg of 17.35 Å, while that of non-halogenated is 16.70 Å. In summary, the MD analyses indicate that halogenation had different effects on protein structural stability: enhancing structural stability in 2ZXV, having little impact in

3VN3, and decreasing structural stability in 5KI8.

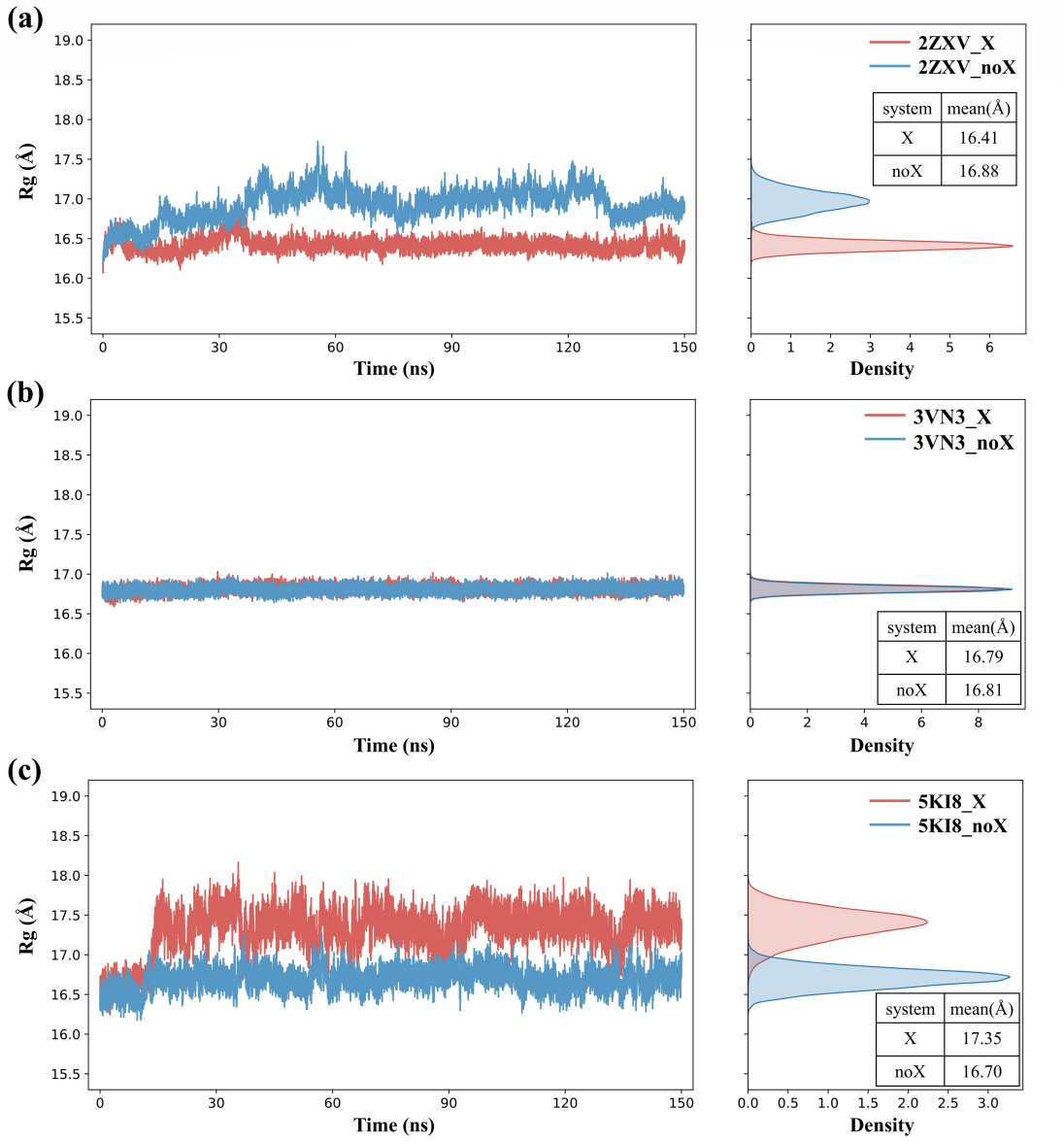

Figure 5 Structural stability of the halogenated (X) and non-halogenated (noX) proteins for (a) 2ZXV, (b) 3VN3 and (c) 5KI8. Left column shows the Rgs of the three systems, while the distribution of Rgs during the stable time period and the mean values were shown in the right column (The common stable time period was selected for the trajectories before and after halogenation based on the time-dependent changes in RMSD as shown in Figure S4). The other two repeated simulation results for each system are shown in Figures S5 and S6.

The MD results across the three systems reveal that introducing halogens yields varying impacts on protein structural stability. To uncover the underlying reasons for these diverse structural stability changes induced by halogens, we conducted further investigations. As protein structural stability is linked to weak intramolecular interactions, the introduction of intra_XBs through halogen incorporation may increase protein structural stability[54]. Therefore, we analyzed the intra_XB formation from the combined trajectories of three systems and found that XBs were detected in 2ZXV_I and 3VN3_I, but not in 5KI8_Br system. The free energy landscapes (FELs) of 2ZXV_I and 3VN3_I trajectories are shown in Figure 6, exhibiting a single free energy minimum centered at relatively short I···O distances and interaction angles of ~146° and ~168° respectively. We further performed QM/MM optimization on the structures of the lowest free energy, and the optimized intra_XB regions formed a better linear contact (Figures 6c and 6f). These suggest that in the 2ZXV and 3VN3 systems, the conformations that form intra_XB are the dominant structures (The details of XBs region for the systems calculated here are provided in Table S8). Furthermore, in 2ZXV_I and 3VN3_I, the halogen is positioned meta to the tyrosine. In 150 ns trajectories, we observed the formation of a certain proportion of intramolecular HBs (with the hydrogen on the hydroxyl group of tyrosine acting as HB donor and halogen acting as HB acceptor), which can enhance the strength of the XBs (Figure S7 and Table S7). As for the 5KI8 system, due to the atoms that can serve as XB acceptors were far away from the halogen (Figure S8), and no intra_XB was detected in the MD trajectories, we believe that there is no intra_XB formed in the 5KI8 system.

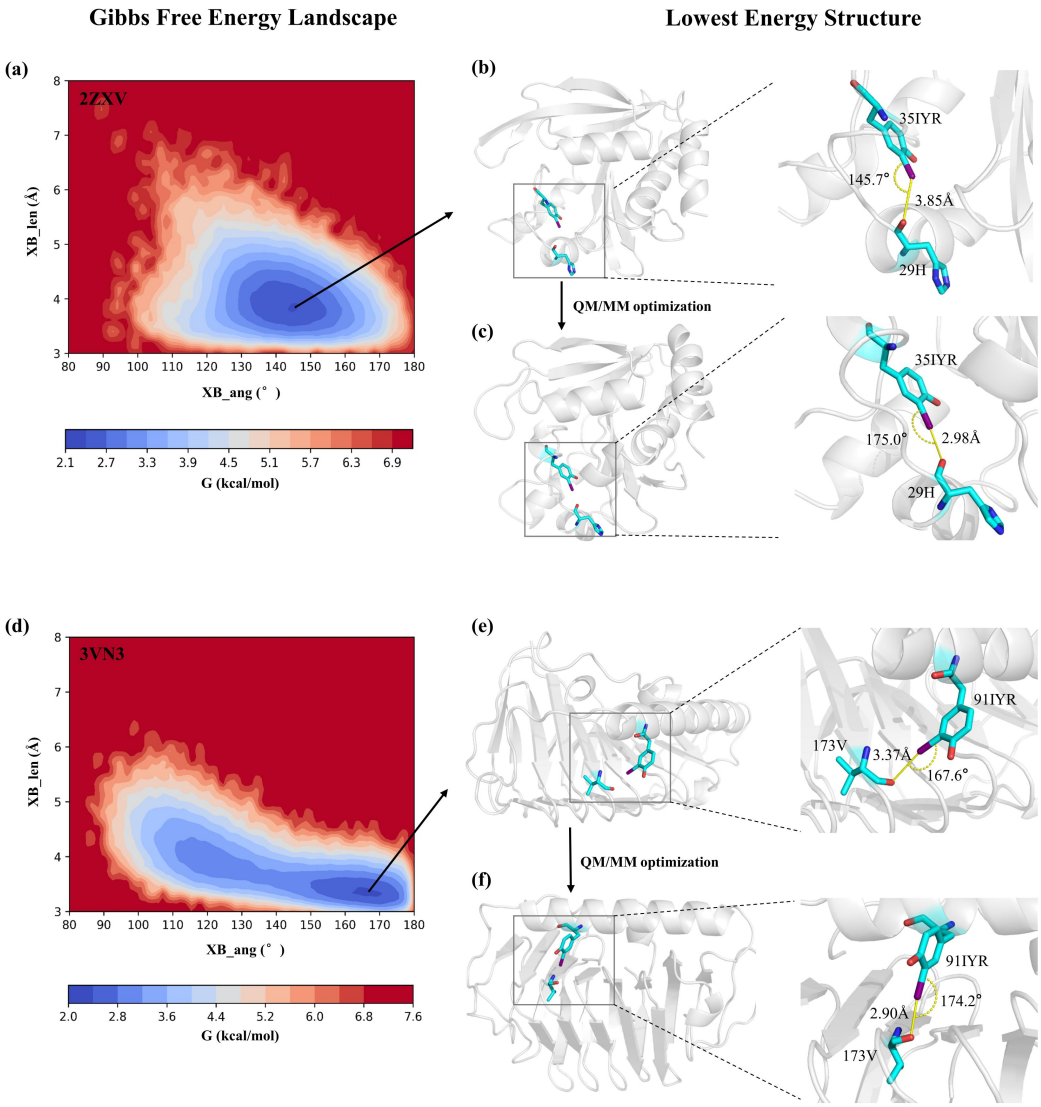

Figure 6 Dominant conformations of 2ZXV_I and 3VN3_I. (a, d) FELs obtained from the simulations of 2ZXV_I and 3VN3_I along with (b, e) the corresponding lowest energy structures at 300 K. FEL was calculated using the bond length and bond angle formed by the atoms related to the intra_XB as the reaction coordinates. The values are given in kcal/mol as indicated by the color bars. (c, f) QM/MM optimization results of the lowest energy structures obtained from the MD trajectories.

We wondered whether the formation of intra_XBs in proteins might also lead to increased other contacts. Therefore, we performed LIT (local interaction time) analysis for each system. Residue with higher LIT value typically demonstrates a

greater degree of contact with surrounding residues. The difference in local interaction time (ΔLIT, LIT_X - LIT_{noX}) was then calculated on the combined trajectories within the stable time period to explore whether protein halogenation affects the degree of other residues' contacts and thus protein stability. In the case of the 2ZXV system (Figure 7a), residues within a 6 Å radius around the halogenated residue ^{35I}TYR (residue 26-42) showed ΔLIT values greater than 0, with a total ΔLIT sum of 12.28%. This suggests that 2ZXV_I has a larger overall LIT compared to 2ZXV_noX, indicating that the introduction of halogen atoms enhances the LIT within the 2ZXV system. For the 3VN3 system, ΔLIT values exhibit minor fluctuations around the 0 value, and when combined with statistical analysis results (Figure 7c), it can be inferred that there is relatively little difference in the overall LIT before and after halogenation. In contrast, in the 5KI8 system, statistical analysis reveals a reversed trend for ΔLIT (Figure 7e), with a total ΔLIT sum of -4.65%. This indicates that the introduction of halogen atoms decreases the LIT within the 5KI8 system. RMSF analyses demonstrated that the introduction of halogen atoms had a noticeable impact on the fluctuations of residues in the 2ZXV and 5KI8 systems (Figures 7b and 7f), while its effect on the 3VN3 system was minimal (Figure 7d), possibly due to the inherent structural rigidity of 3VN3.

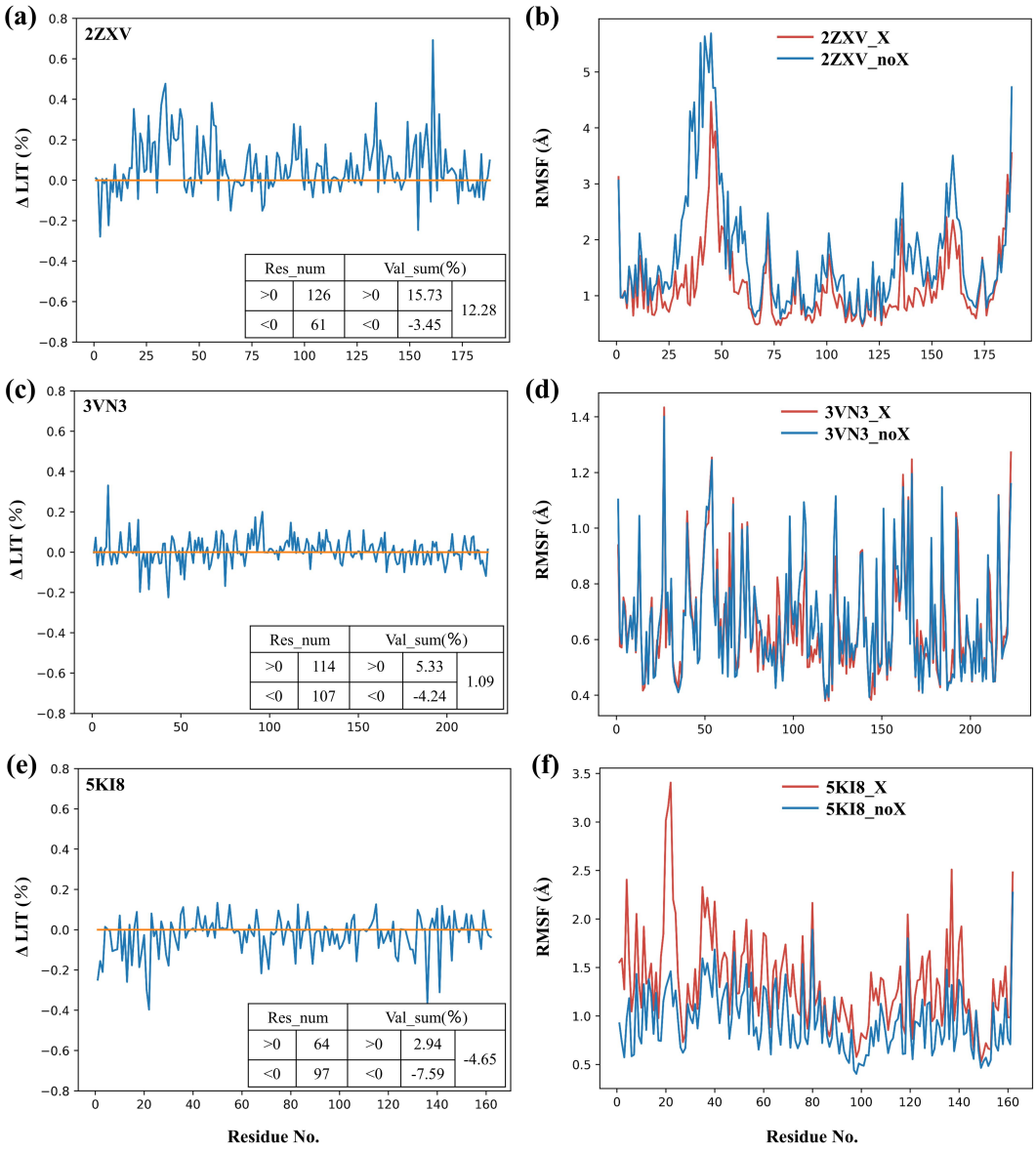

Figure 7 Local characteristics of residues in 2ZXV, 3VN3 and 5KI8. (a, c, e) ΔLIT among the three systems. The number of residues with ΔLIT > 0 or ΔLIT < 0 and the sum of ΔLIT > 0 or ΔLIT < 0 are labeled in the embedded tables. (b, d, f) The RMSF for the Cα atoms among the three systems.

The MD simulation results suggest that the introduction of halogen atoms to form intra_XBs appears to enhance the protein structural stability (as observed in 2ZXV). Conversely, the introduction of halogen atoms without the formation of intra_XBs may negatively impact the formation of residue contacts, leading to a decrease in protein structural stability (5KI8). Additionally, the structural

characteristics of the protein itself may also be a contributing factor. For structurally rigid proteins (such as 3VN3), the structural impact on stability is relatively small even when intra_XBs are formed.

4. Conclusion

In this work, we made a systematic survey on the PDB database for the presence of halogenated residue XBs (hr_XBs) in proteins/peptides and found that the hr_XBs involved in proteins/peptides could be classified into 2 cases: one is the intermolecular halogenated residue XBs (inter_XBs) at the protein-peptide binding interface, the other is the intramolecular halogenated residue XBs (intra_XBs) within the proteins. Statistical results show that the hr_XBs are more prone to utilizing side chain atoms (nitrogen and oxygen) as XB acceptors than main chain atoms, whereas this trend is not notably pronounced in organohalogen XBs. Compared with halogenated ligands, the proteins/peptides have more space to form strong XBs after QM/MM optimization due to the flexibility of residues. Regarding the protein-peptide complexes (5ITF, 1CZI, 1GA4), strong inter_XB interactions are discovered in all three systems and can enhance the binding affinity of the protein-peptide complexes. For the proteins with halogenated residues (2ZXV, 3VN3, 5KI8), halogens have different impacts on protein structural stability, depending on the formation of intra_XBs and the structural characteristics of proteins. In detail, the introduction of halogen atoms to form intra_XBs appears to enhance the protein structural stability. However, the rigidity of the structure can mask the stabilizing effects brought about by the intra_XBs. In comparison, the introduction of halogens that cannot form intra_XBs may reduce the structural stability of the protein. The findings above provide a potential guide for protein engineering and drug design.

Supporting Information

Table S1: Distance criterion of XB length; Table S2: EP Parameters used for MD simulations; Table S3: The calculation results of the optimized structure for 1OKW;

Table S4: The binding energies of optimized structures for 4 systems in chloroform; Table S5: The binding energies of crystal structures for 4 systems in vacuum and chloroform; Table S6: The calculation results of optimized structures of 5ITF; Table S7: Binding energy of XB before and after the introduction of intramolecular HBs; Table S8: The details of the XBs calculated in this study; Figure S1: The propensity of atom types around halogens in proteins/peptides; Figure S2: The analyses of halogenated structures before and after optimization; Figure S3: The analyses of optimized structures before and after halogenation; Figure S4: RMSD analysis of MD trajectories for three Systems; Figure S5: The Rg results of the other two simulation replicates for each system; Figure S6: The RMSD results of the other two simulation replicates for each system; Figure S7: Formation of intramolecular HBs on halogenated Tyr in 2ZXV and 3VN3 systems; Figure S8: Structural details of the 5KI8 system (PDF)

Acknowledgments

The authors sincerely acknowledge the financial support from the National Natural Science Foundation of China (82322067, 82273851) and the National Key Research and Development Program of China (2022YFA1004304).

Data and Software Availability

All data used for calculated were obtained from PDB database. The parameter files (charge, modified force field) required to construct halogenated residues topology files, the python scripts used for protein XBs detection and the initial structures required for calculations in the manuscript are available at the following github repository: https://github.com/Zhijian-Xu/Protein-XB .

ABBREVIATIONS USED

APBS, the adaptive Poisson-Boltzmann equation solver; EP, Extra Point; ESP, electrostatic potential; IGM, Independent Gradient Model; LIT, local interaction time;

MD, molecular dynamics; MEP, molecular electrostatic potential; NBO, Natural Bond Orbital; ONIOM, our own N-layered integrated molecular orbital and molecular mechanics; PDB, Protein Data Bank; PME, particle mesh Ewald method; QM, quantum mechanics; RESP, Restrained Electrostatic Potential; Rg, radius of gyration; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; RRMSE, relative root mean squared error; XB, halogen bond

Author contributions

Conceptualization: Z. X., W. Z.

Methodology: J. L., L. Z., Z. H., L. W., J. Z.

Investigation: J. L.

Visualization: J. L., L. Z.

Supervision: Z. X., W. Z.

Writing—original draft: J. L.

Writing—review & editing: Z. X., W. Z., J. L.

All authors read and approved the final manuscript.

References

1. Hassel, O., Structural Aspects of Interatomic Charge-Transfer Bonding. *Science* **1970,** *170* (3957), 497-&.
2. Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P., Halogen bonding: the sigma-hole. *Journal of Molecular Modeling* **2007,** *13* (2), 291-296.
3. Politzer, P.; Lane, P.; Concha, M. C.; Ma, Y. G.; Murray, J. S., An overview of halogen bonding. *Journal of Molecular Modeling* **2007,** *13* (2), 305-311.
4. Fourmigue, M., Halogen bonding: Recent advances. *Curr Opin Solid St M* **2009,** *13* (3-4), 36-45.
5. Metrangolo, P.; Neukirch, H.; Pilati, T.; Resnati, G., Halogen bonding based recognition processes: A world parallel to hydrogen bonding. *Accounts Chem Res* **2005,** *38* (5), 386-395.
6. Metrangolo, P.; Resnati, G., Chemistry - Halogen versus hydrogen. *Science* **2008,** *321* (5891), 918-919.
7. Riley, K. E.; Hobza, P., Strength and Character of Halogen Bonds in Protein-Ligand Complexes. *Cryst Growth Des* **2011,** *11* (10), 4272-4278.
8. Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M., Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *Journal of Medicinal Chemistry* **2013,** *56* (4), 1363-1388.

9. Mendez, L.; Henriquez, G.; Sirimulla, S.; Narayan, M., Looking Back, Looking Forward at Halogen Bonding in Drug Discovery. *Molecules* **2017,** *22* (9).
10. Scholfield, M. R.; Vander Zanden, C. M.; Carter, M.; Ho, P. S., Halogen bonding (X-bonding): A biological perspective. *Protein Sci* **2013,** *22* (2), 139-152.
11. Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S., Halogen bonds in biological molecules. *P Natl Acad Sci USA* **2004,** *101* (48), 16789-16794.
12. Lu, Y. X.; Shi, T.; Wang, Y.; Yang, H. Y.; Yan, X. H.; Luo, X. M.; Jiang, H. L.; Zhu, W. L., Halogen Bonding-A Novel Interaction for Rational Drug Design? *Journal of Medicinal Chemistry* **2009,** *52* (9), 2854-2862.
13. Hardegger, L. A.; Kuhn, B.; Spinnler, B.; Anselm, L.; Ecabert, R.; Stihle, M.; Gsell, B.; Thoma, R.; Diez, J.; Benz, J.; Plancher, J. M.; Hartmann, G.; Banner, D. W.; Haap, W.; Diederich, F., Systematic Investigation of Halogen Bonding in Protein-Ligand Interactions. *Angew Chem Int Edit* **2011,** *50* (1), 314-318.
14. Wilcken, R.; Zimmermann, M. O.; Lange, A.; Zahn, S.; Boeckler, F. M., Using halogen bonds to address the protein backbone: a systematic evaluation. *Journal of Computer-Aided Molecular Design* **2012,** *26* (8), 935-945.
15. Sirimulla, S.; Bailey, J. B.; Vegesna, R.; Narayan, M., Halogen Interactions in Protein-Ligand Complexes: Implications of Halogen Bonding for Rational Drug Design. *J Chem Inf Model* **2013,** *53* (11), 2781-2791.
16. Riley, K. E.; Murray, J. S.; Fanfrlik, J.; Rezac, J.; Sola, R. J.; Concha, M. C.; Ramos, F. M.; Politzer, P., Halogen bond tunability I: the effects of aromatic fluorine substitution on the strengths of halogen-bonding interactions involving chlorine, bromine, and iodine. *Journal of Molecular Modeling* **2011,** *17* (12), 3309-3318.
17. Parisini, E.; Metrangolo, P.; Pilati, T.; Resnati, G.; Terraneo, G., Halogen bonding in halocarbon-protein complexes: a structural survey. *Chem Soc Rev* **2011,** *40* (5), 2267-2278.
18. Xu, Z.; Yang, Z.; Liu, Y.; Lu, Y.; Chen, K.; Zhu, W., Halogen Bond: Its Role beyond Drug–Target Binding Affinity for Drug Discovery and Development. *Journal of Chemical Information and Modeling* **2014,** *54* (1), 69-78.
19. Pizzi, A.; Pigliacelli, C.; Bergamaschi, G.; Gori, A.; Metrangolo, P., Biomimetic engineering of the molecular recognition and self-assembly of peptides and proteins via halogenation. *Coordin Chem Rev* **2020,** *411*.
20. Egner, U.; Kratzschmar, J.; Kreft, B.; Pohlenz, H. D.; Schneider, M., The target discovery process. *Chembiochem* **2005,** *6* (3), 468-479.
21. Scholfield, M. R.; Ford, M. C.; Carlsson, A. C. C.; Butta, H.; Mehl, R. A.; Ho, P. S., Structure-Energy Relationships of Halogen Bonds in Proteins. *Biochemistry-Us* **2017,** *56* (22), 2794-2802.
22. Carlsson, A. C. C.; Scholfield, M. R.; Rowe, R. K.; Ford, M. C.; Alexander, A. T.; Mehl, R. A.; Ho, P. S., Increasing Enzyme Stability and Activity through Hydrogen Bond-Enhanced Halogen Bonds. *Biochemistry-Us* **2018,** *57* (28), 4135-4147.
23. Jiang, S. Q.; Zhang, L. J.; Cui, D. B.; Yao, Z. Q.; Gao, B.; Lin, J. P.; Wei, D. Z., The Important Role of Halogen Bond in Substrate Selectivity of Enzymatic Catalysis. *Sci Rep-Uk* **2016,** *6*.
24. Danelius, E.; Andersson, H.; Jarvoll, P.; Lood, K.; Grafenstein, J.; Erdelyi, M., Halogen Bonding: A Powerful Tool for Modulation of Peptide Conformation. *Biochemistry-Us* **2017,** *56* (25), 3265-3272.
25. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.;

Bourne, P. E., The Protein Data Bank. *Nucleic Acids Research* **2000,** *28* (1), 235-242.
26. Zhang, Q.; Xu, Z.; Zhu, W., The Underestimated Halogen Bonds Forming with Protein Side Chains in Drug Discovery and Design. *Journal of Chemical Information and Modeling* **2017,** *57* (1), 22-26.
27. Zhang, Q.; Xu, Z. J.; Shi, J. Y.; Zhu, W. L., Underestimated Halogen Bonds Forming with Protein Backbone in Protein Data Bank. *J Chem Inf Model* **2017,** *57* (7), 1529-1534.
28. Xu, Z. J.; Zhang, Q.; Shi, J. Y.; Zhu, W. L., Underestimated Noncovalent Interactions in Protein Data Bank. *J Chem Inf Model* **2019,** *59* (8), 3389-3399.
29. Mu, K. J.; Zhu, Z. D.; Abula, A.; Peng, C.; Zhu, W. L.; Xu, Z. J., Halogen Bonds Exist between Noncovalent Ligands and Natural Nucleic Acids. *J Med Chem* **2022,** *65* (6), 4424-4435.
30. Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J., Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J Chem Theory Comput* **2006,** *2* (3), 815-26.
31. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. B.01*, Wallingford, CT, 2016.
32. Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts* **2008,** *120* (1), 215-241.
33. Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A., 6-31G*basis set for third-row atoms. *Journal of Computational Chemistry* **2001,** *22* (9), 976-984.
34. Wadt, W. R.; Hay, P. J., Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi. *J. Chem. Phys.* **1985,** *82*, 284-298.
35. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *J Am Chem Soc* **1996,** *118* (9), 2309-2309.
36. Lefebvre, C.; Rubez, G.; Khartabil, H.; Boisson, J. C.; Contreras-Garcia, J.; Henon, E., Accurately extracting the signature of intermolecular interactions present in the NCI plot of the reduced density gradient versus electron density. *Physical Chemistry Chemical Physics* **2017,** *19* (27), 17928-17936.
37. Lu, T.; Chen, Q. X., Independent gradient model based on Hirshfeld partition: A new method for visual study of interactions in chemical systems. *Journal of Computational Chemistry* **2022,** *43* (8), 539-555.
38. Lu, T.; Chen, F. W., Multiwfn: A multifunctional wavefunction analyzer. *Journal of Computational Chemistry* **2012,** *33* (5), 580-592.

39. Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A., Self-Consistent Molecular-Orbital Methods .20. Basis Set for Correlated Wave-Functions. *J Chem Phys* **1980,** *72* (1), 650-654.
40. Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J. H.; Liles, K.; Chun, M. J.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W. H.; Krasny, R.; Wei, G. W.; Holst, M. J.; McCammon, J. A.; Baker, N. A., Improvements to the APBS biomolecular solvation software suite. *Protein Sci* **2018,** *27* (1), 112-128.
41. Boys, S. F.; Bernardi, F., The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Molecular Physics* **1970,** *19* (4), 553-566.
42. Ibrahim, M. A. A., Molecular Mechanical Study of Halogen Bonding in Drug Discovery. *Journal of Computational Chemistry* **2011,** *32* (12), 2564-2574.
43. Rendine, S.; Pieraccini, S.; Forni, A.; Sironi, M., Halogen bonding in ligand-receptor systems in the framework of classical force fields. *Physical Chemistry Chemical Physics* **2011,** *13* (43), 19508-19516.
44. Ermak, D. L.; Mccammon, J. A., Brownian Dynamics with Hydrodynamic Interactions. *J Chem Phys* **1978,** *69* (4), 1352-1360.
45. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant-Pressure Molecular-Dynamics Algorithms. *J Chem Phys* **1994,** *101* (5), 4177-4189.
46. Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R., Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *J Chem Phys* **1995,** *103* (11), 4613-4621.
47. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977,** *23* (3), 327-341.
48. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.
49. Mercadante, D.; Grater, F.; Daday, C., CONAN: A Tool to Decode Dynamical Information from Molecular Interaction Maps. *Biophys J* **2018,** *114* (6), 1267-1273.
50. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O., MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **2011,** *32* (10), 2319-2327.
51. Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J.; Melo, M. N.; Seyler, S. L.; Domanski, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M. In *MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations*, Proceedings of the 15th python in science conference, SciPy Austin, TX: 2016; p 105.
52. Zhou, L. P.; Li, J. T.; Shi, Y. L.; Wu, L. Y.; Zhu, W. L.; Xu, Z. J., Preferred microenvironments of halogen bonds and hydrogen bonds revealed using statistics and QM/MM calculation studies. *Physical Chemistry Chemical Physics* **2023,** *25* (26), 17692-17699.
53. Zhu, Z.; Wang, G.; Xu, Z.; Chen, Z.; Wang, J.; Shi, J.; Zhu, W., Halogen bonding in differently charged complexes: basic profile, essential interaction terms and intrinsic σ-hole. *Physical Chemistry Chemical Physics* **2019,** *21* (27), 15106-15119.
54. Newberry, R. W.; Raines, R. T., Secondary Forces in Protein Folding. *Acs Chem Biol* **2019,** *14* (8), 1677-1686.

For Table of Contents Only